[Reprint (with correction of some typographical errors) of pp. 3.1-3.17 in *Proceedings Third International Workshop on Digital Signal Processing Techniques Applied to Space Communications,* ESTEC, Noordwijk, The Netherlands, 23-25 September 1992]

Coding and Modulation for Code-Division Multiple Accessing*

James L. Massey

Signal and Information Processing Laboratory
Swiss Federal Institute of Technology
CH-8092 Zürich

Abstract: It is argued that the "rightful purpose" of a CDMA modulation system is to create a good channel for the coding system. This viewpoint leads to a precise formulation of energy efficiency and a novel formulation of bandwidth occupancy as normalized Shannon bandwidth. Various modulation/coding architectures are formulated and compared. It is shown that it is possible to design systems in such a way that the burden of separating and detecting the users' information sequences is divided between the demodulator and the decoders with resulting reduced complexity but with no penalty in either energy efficiency or bandwidth occupancy.

1 Introduction

Before one begins to study modulation systems and coding systems in detail for code-division multiple accessing (or for any other kind of communications for that matter), it is well to reflect on what the *purposes* of these two systems are. As self-evident as this admonition may seem, we must remark that our experience shows that it is heeded more in the breech than in the practice. The reader is thus asked to bear with a recitation of these purposes that is not at all new but that is prone to be forgotten.

For our purposes, the model of an "L-chip synchronous CDMA channel" shown in Fig. 1 will be adequate. As shown, the model describes one use of a memoryless channel. Each of the M-users, say user i, sends a symbol B_i to his corresponding modulator. We assume that B_i is a random variable with second

*The research reported here was supported by the European Space Research and Technology Centre (ESTEC) in Noordwijk, The Netherlands, under ESTEC Contract No. 8696/89/NL/US

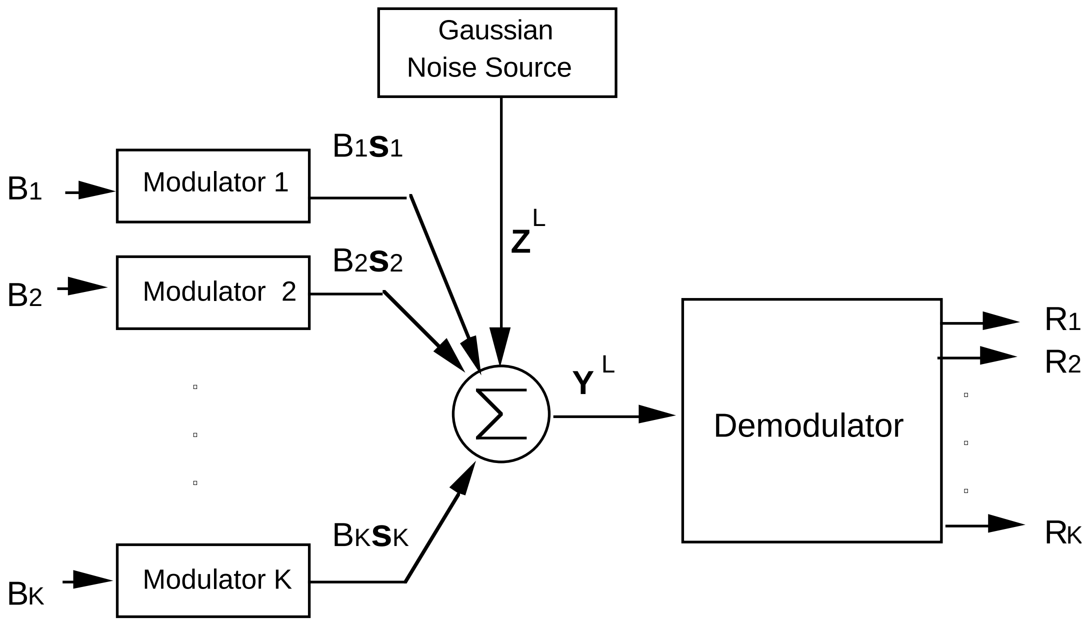

Fig. 1: Model of an L-chip synchronous CDMA channel.
($E[B_i^2] \le E_c$, $\mathbf{s}_i^T \mathbf{s}_i = L$, and $\mathbf{Z}^L$ is white Gaussian noise of variance $N_o/2$.)

moment $E[B_i^2] \le E_c$. We further assume that the random variables $B_1, B_2, ..., B_M$ are independent, which is the essential difference between a multiple-access channel and a single-user channel. Using the usual signal-space representation with respect to any convenient basis of orthonormal waveforms, we can view the modulator as converting B_i into the signal vector $B_i\mathbf{s}_i$, where $\mathbf{s}_i = (s_{i1}, s_{i2}, ..., s_{iL})$ is an L-component vector with energy L. For convenience, we will consider the components of $\mathbf{s}_i$ to be real numbers, i.e., $\mathbf{s}_i \in \Re^L$, which is equivalent to considering a baseband channel (but all our conclusions carry directly over to the passband case when these components are complex numbers). The energy of $\mathbf{s}_i$ is thus

$$\mathbf{s}_i^T\mathbf{s}_i = \sum_{n=1}^{L} s_{in}^2 = L \tag{1}$$

(where the superscript T denotes transposition) and our normalization of this energy to L is motivated by the usual case in CDMA systems where $s_{in} \in \{-1, +1\}$ for all n. The demodulator input is the received vector

$$\mathbf{Y}^L = \sum_{i=1}^{M} B_i \, \mathbf{s}_i + \mathbf{Z}^L \tag{2}$$

where $\mathbf{Z}^L = (Z_1, Z_2, ..., Z_N)$ is *white Gaussian noise* (WGN), i.e., its components are independent zero-mean Gaussian random variables all with the same variance, which as usual we take to be $N_o/2$. We will be less explicit about the nature of the demodulator outputs, since these are at the choice of the designer of the modulation system.

We can now address the question of what the modulation system of Fig. 1 should be designed to do. But this requires us to think more carefully about how the users will actually send information over this channel. In fact, we know that, for reliable and efficient transmission, each user will *encode* his information bits into a sequence of symbols to be sent over the channel, and the receiver will *decode* the corresponding received sequence provided by the demodulator to recover these information bits. The conclusion is inescapable:

> *(i) The rightful purpose of the modulation system in Fig. 1 (i.e., of the K modulators and the demodulator) is to create a good channel for coding.*

To make this general purpose more directly applicable, we must separately consider the modulators and demodulator. The choice of the modulators, i.e., the choice of the sequences $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_L$, determines a multiple-access channel with input $\mathbf{B} = (B_1, B_2, ..., B_M)$ and output $\mathbf{Y}^L = (Y_1, Y_2, ..., Y_L)$. The fundamental measure of goodness of a channel for coding is, of course, its *capacity* (appropriately defined as we will later do) so we must conclude:

> *(ii) The rightful purpose of the K modulators in Fig. 1 is to create a channel from* $\mathbf{B}$ *to* $\mathbf{Y}^L$ *with maximum capacity.*

The question of what the demodulator should do is more subtle. If we continue to take capacity as our only criterion of goodness, the answer is simple: do nothing! Any processing of $\mathbf{Y}^L$ can only reduce capacity since, by the Data Processing Theorem [1, p. 158] of information theory, $I(\mathbf{B}; \mathbf{Y}^L) \geq I(\mathbf{B}; \mathbf{R})$ where $\mathbf{R} = (R_1, R_2, ..., R_M)$ and where I(. ; .) denotes mutual information. Thus, taking $R_i = \mathbf{Y}^L$ for all i (so that we really have only one demodulator output) is optimum in the sense of maximizing capacity. But a channel "good for coding" should mean not only a channel with large capacity but also one that is convenient for coding and decoding. Because (at least roughly) the complexity of decoding grows with the number of quantization levels at the demodulator output, we can conclude the following:

(iii) The rightful purpose of the demodulator in Fig. 1 is to quantize the received vector $\mathbf{Y}^L$ *as coarsely as possible consistent with a tolerably slight decrease of capacity.*

The three conclusions stated above are, as we have already said, not new. They appear quite explicitly in [2, p. 211], [3], [4], [5] and [6], sometimes with *cut-off rate* (which is also an excellent measure of channel quality for coding in single-user channels) used in place of capacity, but this makes no essential difference. In the following sections, we will study modulation and coding for the CDMA channel in light of these rightful purposes of the modulation system. The novelty of much of what follows suggests that although these "rightful purposes" are virtually self-evident, they are usually ignored in the design and analysis of CDMA systems.

The two aspects of the modulation system of Fig. 1 that are the most frequent source of controversy are its *energy efficiency* and its *"bandwidth efficiency"*. We now wish to define these terms in such a way that we can validly compare systems with different values of L and E_c, but of course we would fix K (the number of users) and N_o (the one-sided noise power spectral density) in any such comparison. Energy efficiency is the more easily understood of these two "efficiencies" so we begin with it.

The components s_{in} of the "spreading sequence" $\mathbf{s}_i$ are traditionally called *chips*. Since user i sends L chips with total energy $E\,[B_i\,\mathbf{s}_i^T\,\mathbf{s}_i\,B_i] = L\,E\,[B_i^2] \leq L\,E_c$, where we have made use of (1), we see that E_c is the (maximum allowed) *average chip energy* for user i. We may draw the immediate conclusion:

(iv) When comparing modulation systems of the type in Fig. 1 with the same values of K *and* E_c, *the appropriate measure of energy utilization is the ratio of the per-user capacity per chip of the channel from* $\mathbf{B}$ *to* $\mathbf{Y}^L$ *to the chip energy* E_c.

By the *chip time* T_c of a CDMA system, we mean the time interval allocated to the transmission of one chip over the underlying waveform channel. More precisely, the *chip rate* $1/T_c$ is the (average) number of chips per second produced by a modulator. The chip rate $1/T_c$ is precisely the *Shannon bandwidth* of the modulated signals, i.e., the number of dimensions of signal space used per second, cf. [6]. This Shannon bandwidth is exactly twice the ordinary *Fourier* *bandwidth* when the corresponding chip waveforms are chosen to minimize the Fourier bandwidth. If two modulation systems (possibly with different values of E_c and L) are such that the first has three times the capacity per chip of the second, then (upon the assumption that the two corresponding coding systems operate at the same fraction of capacity) to accomodate the same information transfer rate one will have to send three times as many chips per second for the second system as for the first, i.e., the second system requires three times as much bandwidth. In fact, the *reciprocal* of the total capacity per chip, i.e., the numbers of chips required per bit of total capacity, can and will be defined here as the *normalized Shannon bandwidth*. We need merely to multiply this normalized bandwidth by the desired total information bit rate to obtain the Shannon bandwidth; a further multiplication by one-half yields the Fourier bandwidth (or a close approximation thereto). We are forced to the following conclusion.

> *(v) When comparing modulation systems of the type in Fig. 1 with the same values of* K *and* E_c*, the appropriate measure of bandwidth occupancy is normalized Shannon bandwidth defined as the reciprocal of the total capacity per chip of the channel from* **B** *to* $\mathbf{Y}^L$.

It should now be apparent that the demands for energy efficiency and bandwidth "efficiency" are to some extent conflicting. Because the capacity per chip generally (but not always as we will see in Section 4 below) increases without limit as the chip energy E_c increases, one can usually be as bandwidth "efficient" as one pleases -- if one doesn't care about energy efficiency. "Bandwidth efficiency" is misleading terminology that should be expunged from the vocabulary of communications engineers; instead, one should speak of *bandwidth occupancy*. There is, however, an upper limit on energy utilization (as we will see below) so that it does make good sense to talk about energy efficiency.

2 Trivial Demodulation with Joint Decoding

As noted above, the optimum demodulator, in the sense of maximizing the capacity of the channel from **B** to **R** in Fig. 1, is that which does nothing (expect of course to project the received waveform into its representation $\mathbf{Y}^L$ in signal space. Thus, it is natural to begin our detailed modulation/coding studies with the architecture of Fig. 2 in which the demodulator merely passes $\mathbf{Y}^L$ along to the *joint decoder* that has full responsibility for separating and detecting the sequences of information bits from the K information sources. We now consider the appropriate notion of capacity for the channel from $\mathbf{B}$ to $\mathbf{Y}^L$, which will serve as a useful upper bound on the capacity from $\mathbf{B}$ to $\mathbf{R}$ for less trivial demodulators.

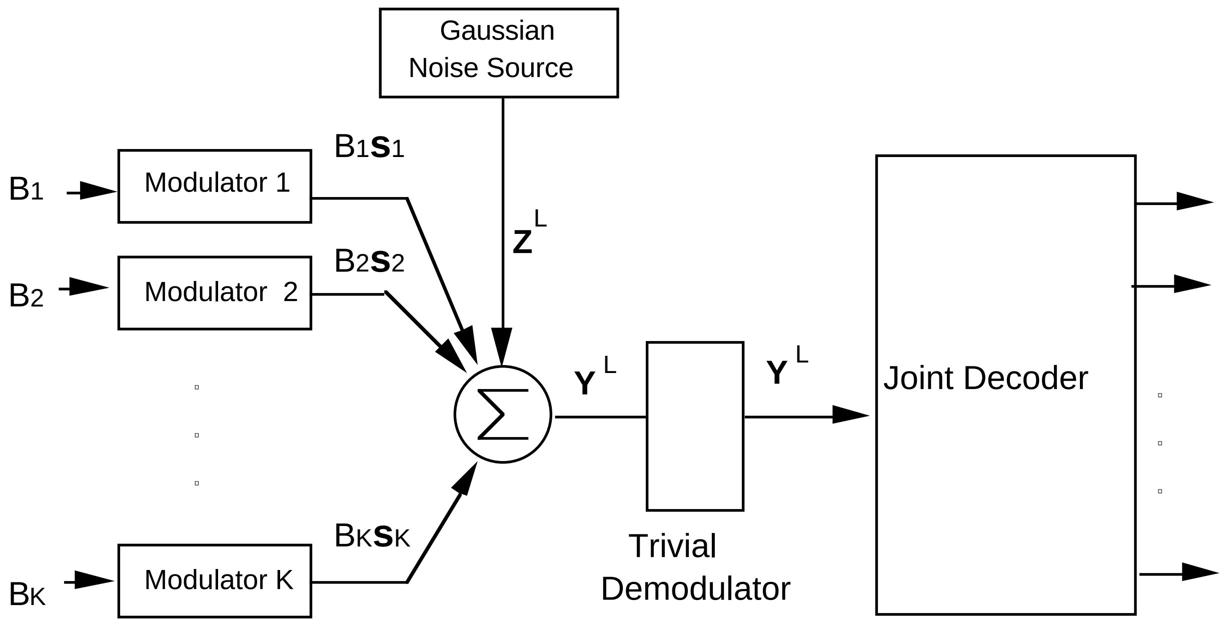

Fig. 2: Trivial demodulation combined with joint decoding--the joint decoders emits decision on the information bits of the K users.

It would seem at first glance that the appropriate notion of capacity for the multiple-access channel would be the *per-user sum capacity per chip*, C_{sum}, from $\mathbf{B}$ to $\mathbf{Y}^L$, i.e., the maximum of $\frac{1}{KL} I(\mathbf{B}; \mathbf{Y}^L)$, where the maximum is taken over all choices of probability distributions for the independent random variables B_1, B_2, ..., B_K subject to the constraint that $E[B_i^2] \le E_c$ for all i. It is easy to obtain a simple upper bound on this capacity. The total average modulated signal energy per chip is at most KE_c with equality if and only if $E[B_i^2] \le E_c$ for all i. The additive Gaussian noise on each chip has variance $N_o/2$. Thus, $\frac{1}{KL} I(\mathbf{B}; \mathbf{Y}^L)$ is upper bounded by $\frac{1}{K}$ times the capacity of a single-user discrete-time Gaussian channel with signal energy KE_c and noise variance $N_o/2$. By one of the oldest results in information theory (cf. [2, p. 147[) this capacity is $\frac{1}{2} \log_2(1 + \frac{KE_c}{N_o/2})$ with equality if and only if $\mathbf{Y}^L$ is WGN with variance $KE_c + N_o/2$.

Proposition 1: *The per-user sum capacity per chip of the channel from* **B** *to* $\mathbf{Y}^L$ *satisfies*

$$C_{sum} \leq \frac{1}{2K} \log_2(1 + \frac{KE_c}{N_o/2}) \text{ [bits/chip]} \quad (3)$$

with equality if and only if $\mathbf{Y}^L$ *is WGN with variance* $KE_c + N_o/2$.

But in fact the sum capacity is not the capacity of most interest for multiple-accessing since it places no demands on fairness. What is of much greater interest is the *per-user symmetric capacity per chip*, C_{sym}, which is defined as the maximum number C such that the K-tuple (C, C, .., C) lies in the capacity region of the channel, i.e., such that all of the K users can simultaneously send information reliably at any per-user rate less than C bits/chip. The following somewhat surprising result is proved in [7].

Proposition 2: *The per-user symmetric capacity per chip of the channel from* **B** *to* $\mathbf{Y}^L$ *satisfies*

$$C_{sym} \leq \frac{1}{2K} \log_2(1 + \frac{KE_c}{N_o/2}) \quad \text{[bits/chip]} \quad (4)$$

with equality in the (usual) case where $K \geq L$ *if and only if the* $L \times K$ *matrix* S *with columns* $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$ *has mutually orthogonal and equal-energy rows.*

We will call sequences $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$ satisfying the conditions for equality in Proposition 2 a *Welch-Bound-Equality (WBE) sequence set* (more precise terminology would be "sequence multiset" as the sequences $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$ need not be distinct), since these are precisely the conditions that were proved in [8] for equality to hold in a lower bound due to Welch [9] on the sum of the "squared correlations" $(\mathbf{s}_i^T \mathbf{s}_j)^2$ over all i and j. Because $C_{sym} \leq C_{sum}$, we see from (4) and (6) that *fairness can be achieved with no loss of capacity by choosing the spreading sequences* $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$ *to form a WBE sequence set.*

Because $\log_2(1 + \frac{KE_c}{N_o/2}) \leq 1.44 \frac{KE_c}{N_o/2}$ with equality when and only when $\frac{KE_c}{N_o/2}$ is sufficiently small (say, $\frac{KE_c}{N_o/2} \leq 1$), we obtain the following from (4).

Proposition 3: *The per-user symmetric capacity per chip of the channel from* **B** *to* $\mathbf{Y}^L$ *satisfies*

$$C_{sym} \leq 1.44 \frac{E_c}{N_o} \tag{5}$$

with near equality in the case $K \geq L$ *when and only when* $\frac{E_c}{N_o}$ *is sufficiently small (say,* $\frac{E_c}{N_o} \leq \frac{1}{2K}$*) and the spreading sequences form a WBE sequence set.*

It is now natural [and in keeping with conclusion (iv) of Section 1] to define the *energy efficiency* γ of a CDMA modulation system by

$$\gamma = C_{sym} / (1.44 \frac{E_c}{N_o}) \tag{6}$$

since this percentage cannot exceed 1 but approaches 1 when $\frac{E_c}{N_o} \leq \frac{1}{2K}$ and a WBE sequence set is used.

3 Full Demodulation with Single-User Decoding

The main drawback of the modulation/coding architecture of Fig. 2 is that it places enormous demands on the "joint decoder." Decoding on a single-user channel is generally complex enough; joint decoding for a K-user channel with, say, $K \approx 10^4$ can be mind-boggling in complexity. For this reason, most past CDMA systems use the architecture of Fig. 3, in which by a *single-user decoder for user i is meant a decoder that treats the contributions of all other users to its input as constituting additional white Gaussian noise.* The demodulator in Fig. 2 has been given full responsibility for separating the signals of the K users!

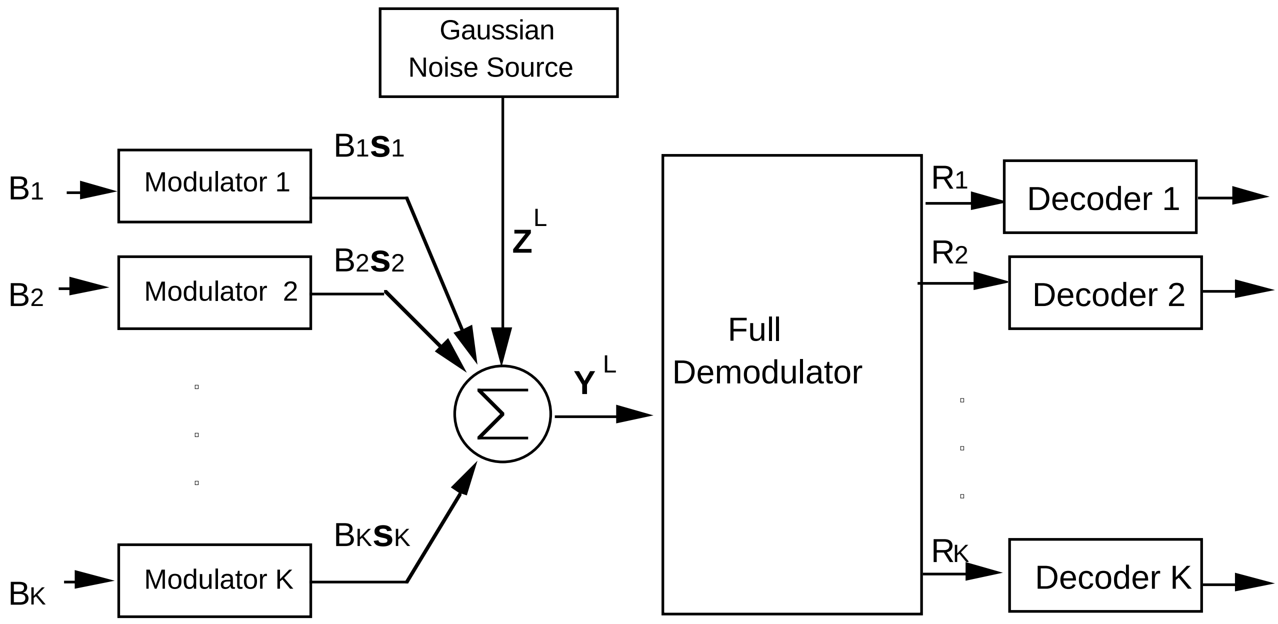

Fig. 3: Full demodulation with single-user decoding wherein the decoder for user i treats the effects of all other users on R_i as white Gaussian noise.

Consider without loss of essential generality the situation for user 1, and assume here and hereafter that $E[B_i] = 0$ and that $E[B_i^2] = E_c$ for all i, which conditions entail no loss of optimality and are generally desirable in practice. The optimum demodulator will, of course, supply to the single-user decoder for user 1 the number R_1 determined by correlating $\mathbf{Y}^L$ with user 1's spreading sequence $\mathbf{s}_1$, i.e.,

$$R_1 = \frac{1}{\sqrt{L}} \mathbf{s}_1^T \mathbf{Y}^L = \sqrt{L}\, B_1 + W_1 + N_1 \tag{7}$$

where

$$N_1 = \frac{1}{\sqrt{L}} \mathbf{s}_1^T \mathbf{Z}^T$$

is a zero-mean Gaussian random variable with variance $N_o/2$ and where

$$W_1 = \frac{1}{\sqrt{L}} \sum_{j=2}^{K} B_j\, \mathbf{s}_1^T \mathbf{s}_j \tag{8}$$

is a zero-mean random variable with variance

$$\frac{E_c}{L} \sum_{j=2}^{K} (\mathbf{s}_1^T \mathbf{s}_j)^2.$$

By the central-limit theorem (cf. [10, p. 256]), one sees from (8) that W_1 will be well approximated by a Gaussian random variable for a WBE sequence set and, indeed, most other sequence sets of practical interest. But Welch's bound [9] effectively provides a lower bound on the sum of the variances of the random variables $W_1, W_2, \ldots, W_K$ that was shown in [8] to hold with equality if and only if the sequence set is WBE. It was further proved in [8] that, for a WBE sequence set, these variances all have the same value, namely $E_c(K - L)$. Thus, we see that (7) describes a single-user Gaussian channel from B_i to R_i with signal energy L Ec and noise variance at least $E_c(K - L) + N_o/2$ where the minimum holds if and only if the sequence set is WBE. Noting that user i uses this equivalent channel only once every L chips, we may summarize our observations as follows.

Proposition 4: *The per-user single-user-decoding capacity per chip of the channel from* **B** to **R** *for the (usual) case $K \geq L$ satisfies*

$$C_{sud} \leq \frac{1}{2L} \log_2[1 + \frac{L\,E_c}{N_o/2 + E_c(K - L)}] \quad \text{[bits/chip]} \qquad (9)$$

with equality if and only if $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$ *form a WBE sequence set. Moreover, if* $K \geq 2L$ *or if* $N_o/2 \geq E_c(K - L)$, then C_{sud} is well approximated for a WBE sequence set by

$$C_{sud} \approx 0.72 \frac{E_c}{N_o/2 + E_c(K - L)} \quad \text{[bits/chip].} \qquad (10)$$

The first of the two conditions for the validity of the approximation (10) follows from the fact that if $K \geq 2L$, then the second term in square brackets in (9) is less than 1 so that the logarithm is well approximated by $\log_2(e) = 1.44$ times that term.

We see from (9) that the approximation

$$C_{sud} \approx 1.44 \frac{E_c}{N_o} \quad \text{[bits/chip]} \qquad (11)$$

holds when and only when the sequence set is WBE and, moreover, $E_c(K - L)$ is sufficiently small compared to $N_o/2$, say $\frac{E_c}{N_o} \leq \frac{1}{4(K - L)}$. But (11) is precisely the condition for near 100% energy efficiency, i.e., for $\gamma = 1$ in (6). Note also that the condition $\frac{E_c}{N_o} \leq \frac{1}{4(K - L)}$ is only about twice as restrictive (since $K \geq 2L$ is assumed and usually $K >> L$) as the condition $\frac{E_c}{N_o} \leq \frac{1}{2K}$ required for near 100% energy efficiency with joint decoding. We can conclude the following.

Both a well-designed single-user-decoding system and a joint-decoding system can operate as close to 100% energy efficiency as desired by choosing a suitably small "signal-to-noise ratio" $\frac{E_c}{N_o}$. *Moreover, the upper limit of "signal-to-noise ratios"* $\frac{E_c}{N_o}$ *for which near 100% energy efficiency is possible is only about twice as large for joint decoding as for single-user decoding.*

4 Bandwidth Comparisons

As we have just seen, suboptimum (but simple) single-user decoding implies no unavoidable penalty in energy efficiency compared to optimum (but complex) joint decoding. We now show, however, that single-user decoding does entail a substantial bandwidth penalty compared to joint decoding.

Consider first single-user decoding. To operate at energy efficiency γ requires, by the definition (6), that

$$C_{sud} = 1.44\,\gamma \frac{E_c}{N_o} \quad \text{[bit/chip]}. \tag{12}$$

Equating the right sides of (10) and (12), then solving for the "signal-to-noise ratio" $\frac{E_c}{N_o}$ yields

$$\frac{E_c}{N_o} = \frac{1 - \gamma}{\gamma (1 - L/K)\, 2\, K} .$$

Substituting this into (12) gives the following expression for the normalized Shannon bandwidth that, like (10), is a close approximation whenever $K \geq 2L$.

$$(K\,C_{sud})^{-1} \approx (1 - L/K)\,\frac{1.39}{1-\gamma} \quad \text{[chips/bit]}. \tag{13}$$

The situation for joint decoding is somewhat more complicated. Again, we begin with

$$C_{sym} = 1.44\,\gamma\,\frac{E_c}{N_o} \quad \text{[bits/chip]} \tag{14}$$

and then equate the right sides of (4) and (10) to obtain the following transcendental equation for the corresponding "signal to-noise" ratio:

$$2\,\gamma\,K\,\frac{E_c}{N_o} = \mathit{ln}(1 + 2K\frac{E_c}{N_o}). \tag{15}$$

When E_c is sufficiently small (which implies $\gamma \approx 1$), one finds from (15) that

$$\frac{E_c}{N_o} \approx \frac{1-\gamma}{K},$$

which when used with (14) yields the following approximation for the normalized Shannon bandwidth:

$$(K\,C_{sym})^{-1} \approx \frac{0.69}{1-\gamma}, \tag{16}$$

which holds when $K \geq 2L$ and $\gamma \approx 1$. Comparing (13) and (16) shows that, in the usual case when $K >> L$,

> *the bandwidth required to operate at a given high energy efficiency* γ *is about twice as great for a single-user-decoding system as for a joint-decoding system.*

But high energy efficiency is in fact the most favorable case for the bandwidth required with single-user decoding compared to that for joint decoding, as one sees from Table I where we have listed the required normalized Shannon bandwidths for single-user decoding [obtained from (13)] and for joint decoding [obtained by solving the transcendental equation (15)]. At an energy efficiency of 20%, which is in the region where one could expect a practical CDMA system to operate, one sees that a single-user-decoding system will require well more than three times the bandwidth of a joint-decoding system in the usual case where K >> L. One is tempted to speculate that the poor "spectral efficiencies" that have sometimes been reported for CDMA systems are rather the result of their designers' predilection for single-user processing rather than of any inherent spectral disadvantage of CDMA systems compared to other multiple-access systems (there is no such disadvantage!).

γ	$(K\,C_{sud})^{-1}$ [chips/bit]	$(K\,C_{sym})^{-1}$ [chips/bit]
1	∞	∞
.9	13.9 (1 - L/K)	6.71
.8	6.95 (1 - L/K)	3.23
.7	4.63 (1 - L/K)	2.06
.6	3.48 (1 - L/K)	1.47
.5	2.78 (1 - L/K)	1.11
.4	2.32 (1 - L/K)	0.86
.3	1.99 (1 - L/K)	0.67
.2	1.74 (1 - L/K)	0.52
.1	1.54 (1 - L/K)	0.38
.01	1.40 (1 - L/K)	0.21
.001	1.39 (1 - L/K)	0.15
≈ 0	1.39 (1 - L/K)	≈ 0

Table I: Normalized Shannon bandwidths $(K\,C_{sud})^{-1}$ and $(K\,C_{sym})^{-1}$ for single-user decoding (with $K \geq 2L$ assumed) and joint-decoding CDMA systems, respectively, as a function of energy efficiency γ.

One sees also from Table I that one cannot reduce the normalized Shannon bandwidth of the single-user-decoding system below 1.39 (1 - L/K), no matter how inefficiently one is willing to use energy. This is a result of the fact that, as follows from (9),

$$KC_{sud} \rightarrow \frac{K}{2L} \log_2 \frac{1}{1 - L/K} \quad \text{[bits/chip]} \quad \text{as } E_c \rightarrow \infty,$$

which implies when K >> L that

$$KC_{sud} \le 0.72 \quad \text{[bits/chip]}$$

with near equality for large E_c. In other words, when K >> L, no amount of chip energy can give a total capacity of more than 0.72 [bits/chip] if single-user decoding is used. Hence, the normalized Shannon bandwidth cannot be reduced below 1/0.72 = 1.39 [chips/bit]. For joint decoding, however, one sees from (4) that

$$KC_{sym} \to \infty \quad \text{as} \quad E_c \to \infty,$$

which is why the normalized Shannon bandwidth can be made as small as desired if one is will to live with small energy efficiencies.

5 Partial Demodulation with User-Group Decoding

As we will now show, it is possible to make a compromise between single-user decoding and joint decoding that combines the best features of both, namely, the reduced complexity of the former and the reduced bandwidth of the latter. [The ideas in this section represent joint work of the author and his colleague, Dr. Thomas Mittelholzer.] The "hybrid" architecture is shown in Fig. 4.

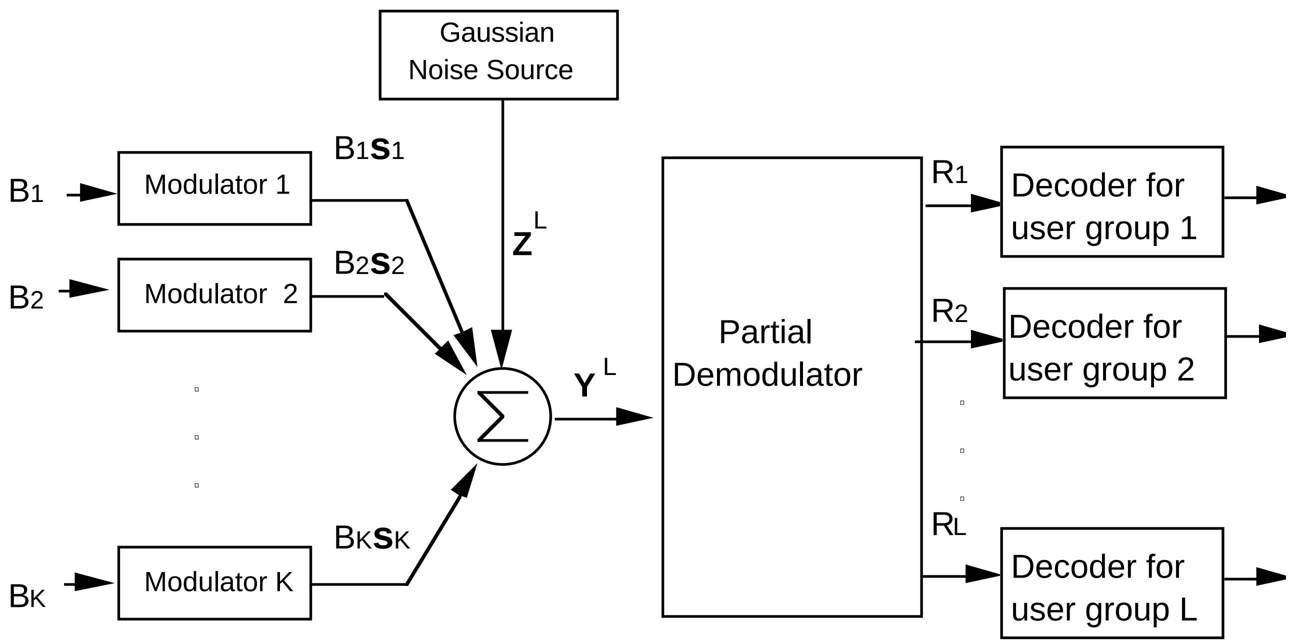

Fig. 4: Partial demodulation into signals for L groups of m users (K = mL) followed by decoding for each group of users.

We assume that K = mL for some positive integer m and that the users are partitioned into L groups of m users each. We further assume that $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L$ are *mutually orthogonal sequences* (i.e., $\mathbf{s}_i^T \mathbf{s}_j = 0$ for $i \neq j$), each of energy L, and that all users in the i-th user group G_i use the same spreading sequence $\mathbf{s}_i$. The $L \times K = L \times mL$ *sequence matrix* S described in Proposition 2 can then be taken as

$$S = [M : M : \ldots : M] \tag{17}$$

where M is the matrix whose columns are $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L$. By hypothesis,

$$M^T M = L I_L$$

where I_L is the $L \times L$ identity matrix. But M^T is then (within a constant factor) the inverse of M so that

$$M M^T = L I_L \tag{18}$$

also holds, i.e., the *rows* of M are also mutually orthogonal and each of energy L. It follows now from (17) that the rows of S are also mutually orthogonal and each of energy mL = K. Thus, the sequence set for the K = mL users is in fact a WBE sequence set and hence, by Proposition 2, delivers a per-user symmetric capacity for the channel from $\mathbf{B}$ to $\mathbf{Y}^L$ satisfying

$$C_{sym} = \frac{1}{2K} \log_2\left(1 + \frac{KE_c}{N_o/2}\right) \quad \text{[bits/chip]}, \tag{19}$$

which is the best possible.

Now, suppose that the "partial demodulator" of Fig. 4 correlates each of the L sequences assigned to the user groups with the received sequence $\mathbf{Y}^L$ (which requires only L = K/m correlations compared to K correlations for the full demodulator considered in Section 2.) This partial demodulator then delivers to the decoder for the i-th user group the output

$$R_i = \frac{1}{\sqrt{L}} \mathbf{s}_i^T \mathbf{Y}^L = \sqrt{L} \sum_{j \in G_j} B_j + N_i \tag{20}$$

where we have used (2) and the mutual orthogonality of $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L$ and where

$$N_i = \frac{1}{\sqrt{L}} \mathbf{s}_i^T \mathbf{Z}^L$$

is a zero-mean Gaussian random variable of variance $N_o/2$. We can recover $\mathbf{Y}^L$ from $\mathbf{R} = (R_1, R_2, \ldots, R_L)$ in the manner

$$\sum_{j=1}^{L} \sum_{i=1}^{L} \frac{1}{\sqrt{L}} \mathbf{s}_j R_i = \sum_{j=1}^{L} \sum_{i=1}^{L} \frac{1}{L} \mathbf{s}_j \mathbf{s}_i^T \mathbf{Y}^L \qquad (21)$$

$$= \frac{1}{L} M M^T \mathbf{Y}^L = \mathbf{Y}^L$$

where we have made use of (18). This means that the capacity of the channel from $\mathbf{B}$ to $\mathbf{Y}^L$ is exactly the same as that from $\mathbf{B}$ to $\mathbf{R}$. Moreover, we note that the Gaussian random variables $N_1, N_2, \ldots, N_L$ are independent. It follows then from (20) that the inputs Bj for $j \in G_i$ are independent of all demodulator outputs except R_i. Hence, the per-user symmetric capacity C_i of the channel from these inputs to R_i must be precisely the same as the per-user symmetric capacity from $\mathbf{B}$ to $\mathbf{R}$, i.e.,

$$C_i = \frac{1}{2K} \log_2(1 + \frac{KE_c}{N_o/2}) \quad \text{[bits/chip]} \qquad (22)$$

so that no capacity loss has resulted from our partial demodulation. The decoder for the i-th user group can now operate on the sequence of symbols from the single demodulator output labelled R_i to separate and detect the users only in this group and without having to worry at all about the other users. Absolutely no penalty in either energy efficiency or bandwidth occupancy compared to optimum joint decoding is incurred!

Many variations on the above hybrid scheme are possible, but the practical advantage of dividing the task of separating and detecting all M users between the modulation system and the coding system should be sufficiently clear to the reader from the simple scheme that we have described above

6 Concluding Remarks

The reader will certainly have noticed that in our analyses above we have suppressed many important practical aspects of CDMA systems such as lack of synchronization, different user energies, time variations, multipath effects, and the like. This was not because including these details would have altered our conclusions (they would not!) but rather because their inclusion would obscure the conceptual lessons that we were determined to set forth as nakedly as possible. We do not mean to suggest that treating these details is easy. However, unless we keep our eyes clearly fixed on the "rightful purposes" of a CDMA system and on the basic information theory of such systems, we are only too apt to lose sight of the forest because of the many trees of detail.

References

[1] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, Mass.: Addison-Wesley, 1987.

[2] R. M. Fano, *Transmission of Information*. Cambridge, Mass.: MIT Press & Wiley, 1961.

[3] J. M. Wozencraft and R. S. Kennedy, "Modulation and Demodulation for Probabilistic Coding," *IEEE Trans. Info. Th.*, vol. IT-12, pp. 291-297, July 1966.

[4] J. L. Massey, "Coding and Modulation in Digital Communications," *Proc. Int. Zurich Seminar on Digital Comm.*, pp. E2(1)-E2(4), 1974.

[5] J. L. Massey, "The How and Why of Channel Coding," *Proc. Int. Zurich Seminar on Digital Comm.*, pp. 67-73, 1984.

[6] J. L. Massey, "Deep-Space Communications and Coding: A Marriage Made in Heaven," to appear in *Proceedings of the 1992 DLR Seminar* "Advanced Methods for Satellite and Deep Space Comm.", Bonn, 22-23 Sept. 1992.

[7] M. Rupf and J. L. Massey, "Optimum Sequence Multisets for Symbol-Synchronous Code-Division Multiple-Access Channels," submitted to *IEEE Trans. Info. Th.*, 1992.

[8] J. L. Massey and T. Mittelholzer, "Welch's Bound and Sequence Sets for Code-Division Multiple-Access Systems," to appear in *Proc. of Sequences 91: Methods in Security and Computer Science*, Positano, Italy, 17-22 June 1991.

[9] L. R. Welch, "Lower Bounds on the Maximum Cross Correlation of Signals," *IEEE Trans. Info. Th.*, vol. IT-20, pp. 397-399, May 1974.

[10] W. Feller, *An Introduction to Probability Theory and its Applications*. New York: Wiley, 1966.